International Journal of Research in Advent Technology, Vol.6, No.7, July 2018
E-ISSN: 2321-9637
Available online at www.ijrat.org

Pectoral Muscle Identification and Detection of Cancer in Mammograms Using Unsupervised Learning Algorithms

Dr. S. Julian Savari Antony
Department of Electronics and Communication Engineering,
Keshav Memorial Institute of Technology, Hyderabad, India
Email: savari.sm@gmail.com

Abstract- Breast cancer is a severe public health problem in several countries. It presents a well-organized computer-aided mass classification method in mammograms using K Means and Neural Network, which executes benign-malignant classification on the region of interest (ROI) that contains mass. This paper presents a research on mammography images using K-means algorithm for detecting cancer tumor mass and microcalcification to help of initiated centroid values which have separated for a different area of values. One of the major mammographic characteristics for mass classification is texture. Neural Network exploits this important factor to classify the mass into benign or malignant. The statistical textural features used in illustrating the masses is pectoral muscle. It has used MIAS with more number of samples. The foremost aim of the method is to increase the usefulness and efficiency of the classification process in an objective means to diminish the numbers of false-positive of malignancies. The proposed technique shows better results in 0.017799 seconds time complexity.

Index Terms- Mammographic Images, Neural network (NN), K means, Region of Interest, Pectoral Muscle Identification, Computer-aided detection.

1. INTRODUCTION

Breast cancer is one of the most overwhelming reasons of the demise among women in the world and mammography image is quiet the most commonly used method for detecting breast cancer at early stage. However, radiologists can miss a significant portion of abnormalities. Some studies indicate that Computer Aided Detection systems (CADe) can deliver a second opinion to the radiologists and potentially decrease the missed detection rate [1]. A CADe system used in breast cancer screening programs is collected by two main steps: the identification of suspicious regions and the false positives reduction [2]. Algorithms for the False Positive Reduction (FPR) of suspicious signs of disease, can work either with one view or with multiple views [3]. Typically, the one view FPR is a two class's classification task in which each Region of Interest (ROI) can be classified as a mass or as normal breast tissue. A set of geometric and/or textural features have to be extracted and selected to train the classifier. Alternatively, template matching approaches can be used, comparing each extracted ROI with all the ROIs of a certain database using similarity measures or features vectors.

Neural networks have supported for breast cancer detection by several re-searchers. Various efforts to refine classification presentation have made, by a number of plans involving some means of choice between alternatives. Ensembles have been proposed as a mechanism for improving the classification accuracy of existing classifiers [4] providing that elements are diverse.

It designed a CAD system for breast mass detection on digital breast to mosynthesis (DBT) mammograms. Each mass candidate was segmented from the structured background, and its image features were extracted. A feature classifier was designed to differentiate true masses from normal tissues. The CAD system achieved a sensitivity of 85%, with 2.2 false-positive objects per case [5]

It proposed and evaluated the performance of a CAD algorithm in marking preoperative masses. First, Digitized mammograms were processed with an adaptive enhancement filter shadowed by a local border refinement stage. Test results showed that malignant masses were detected with the computer in 87% (135 of 156), 83% (130 of 156), and 77% (120 of 156) of the malignant cases at FPI rates of 1.5, 1.0, and 0.5 marks per mammogram, respectively [6]

They used a recursive median filtering technique that could be applied to images at a number of scales and orientations, giving a scale space description at pixel level. This technique was applied to mammography, in the detection of mass-like structures associated with speculated lesions. A sensitivity of 80% was achieved with 0.25 false positives per image [7]

Classification was another most important process in CAD system design. In [8] used improved local binary pattern operator for mass classification. The papers [9], [10], [11] used support vector machine with combination of different techniques for the classification of masses. Naïve bayes classifier [12], K means classifier, fuzzy C means clustering [13][14] are some of the common methods were used by the previous work. In [15] designed least square support vector machine which provided effective classification compared to other methods. An ANN was configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons [16]. They were used several methods and got from output in different kinds of accuracy, false positives and time complexity.

2. PROPOSED METHOD

The proposed system is following the process as k-means and neural network.

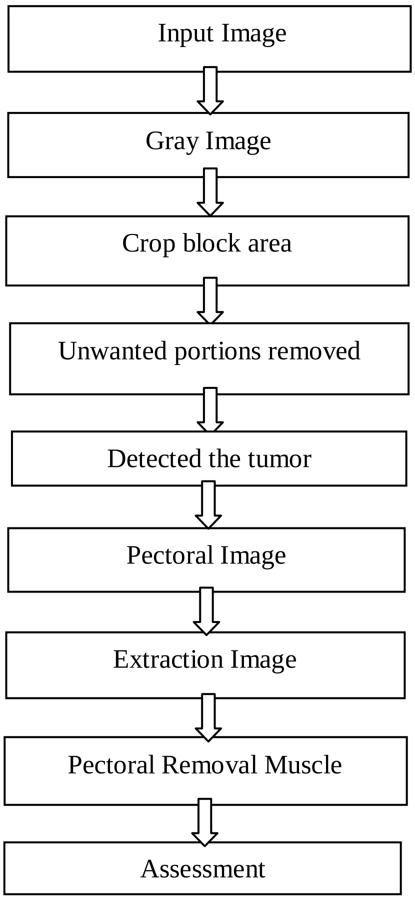


Fig.1. Flow chart showing the Main Steps in K-Means Algorithm and Neural Network

2.1. *Image Pre-processing*

Image pre-processing is a technique to analyze the image before start executing our process. In this technique by using the crop method the size of the original image is re-sized in the selected or exact area. For example, the unwanted area of the real image is eliminated over here in shown fig.2

2.2. *Image Enhancement*

Next step to the image preprocessing is image enhancement. Image enhancement is the process by which we can able to improve the quality of digitally stored images. In enhancement method was improving the quality and value of the original image which is most important for further process in shown fig.2. which is mentioned by UPR.

2.3. *K-Means Algorithm*

Images were classified as cancerous or non-cancerous by best fit into a cluster and are assigned to that cluster. The K-Means algorithms were used for the purpose. The algorithm uses random seeds, i.e., points with random mean values to form lines to separate the classes. Next, the points within the delineated areas are analyzed, and their mean values are calculated. The means form the new seeds from which a new series of lines can be formed to separate the classes. This process is done repeatedly. The advantage of this method is that it has the potential to model complex target functions with a small set of features. The clustering works based on the following equations:

$$D(i,k) = \|(Xk - Vi)\|2 \quad \text{for } I \leq C,\ K \leq N, \qquad (1)$$

$$Vi(l) = (\sum Ni\ Xi\ )/Ni \qquad (2)$$

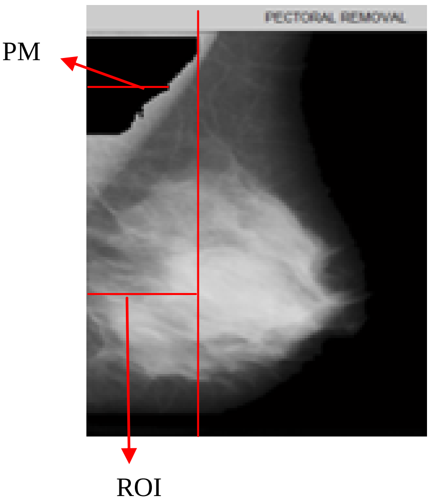
$$\prod_{k=1}^{n} max \,|\, V(l) - V(l\text{-}1) \,|\, \neq 0 \qquad (3)$$

D(i,k) calculates the distances between each class, c is the number of clusters, N is the number of objects in the cluster and v determines the cluster center. The higher value of k results in smooth grouping. The method follows the usual steps to satisfy the primary objective: clustering all the image objects into K distinct groups. First, K centroids are defined, one for each group, being their initial position very important to the result. After that, it is determined a property region for each centroid, which groups a set of similar objects. The interactive stage of the algorithm is started, in which the centroid of each group is recalculated in order to minimize the objective function. This function, for K-means, is the minimum square method, calculated by tumor area. It has divided to centroid four values. The input image applied to k-means that have generated by initiated centroid value each images. To find the tumor area is using equation .3.

The K- Means Algorithm [18] can be done by the following work flow. As a base step an image is taken from a particular portion as a gray image. Crop the block portion of gray image and remove the unwanted portion then we can easily able to detect the tumor portion exactly in fig.2.

Fig.2. Detected the tumor using K-Means algorithm

2.4. *Pectoral Muscle*



PM-Pectoral Muscle; ROI-Region of Interest

Fig.3. Pectoral Removal Muscle

It is preprocessing steps applied to the mammogram is the removal of pectoral muscle, as its presence within the mammogram may adversely affect the outcome of cancer detection processes. The left side unexposed X-ray portion of the mammogram was completely removed such that the top leftmost pixel is a pectoral muscle pixel. Also, the skin–air boundary was roughly strongminded in a few upper rows of the mammogram and was utilized to select the region as ROI which includes the complete pectoral muscle.

2.5. *Neural Network*

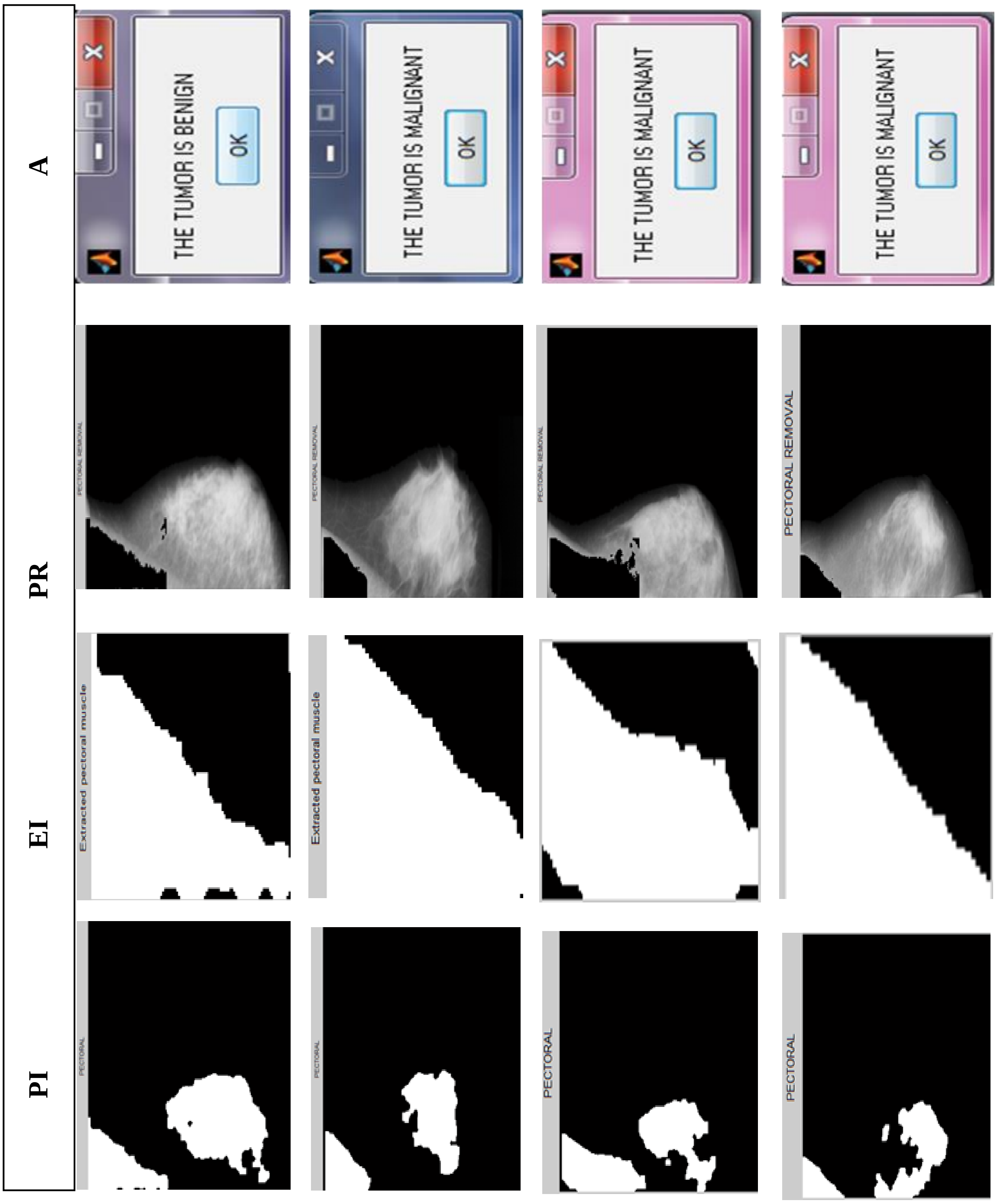
Any classification method uses a set of features or parameters to characterize each object, here these features should be relevant to the task at hand. There are two phases of constructing a classifier. First is the training phase, in which a training set is used to determine how the features are to be weighted and combined in order to classify the objects. Secondly, in the application phase, the weights obtained from the training set are applied to a set of new objects for classification. To obtain a better classification rating, a classifier based on neural networks was designed. The architecture of the network) is a multi-layered one where the nodes in a layer are fully connected to the nodes in the next layer. The input layer contains the fractal feature values, such as fractal dimension and fractal signature. The hidden layer contains five nodes and the output layer has an output node. This neural network is trained using the back propagation algorithm.
Once the tumor is detected then it can be leveled by using neural network. Leveling is the process to determine the effected level of tumor. Once the pectoral image is extracted as the extraction image which is used to remove the pectoral muscle and to determine the tumor as in evil or in beginning stage. A set of twelve images are obtained and neural network operation is made. This shows the Sample Feature Values for Mammogram images. The fig.4 shows that initially, the pectoral image is obtained. Then extraction is carried out in order to obtain the Extraction Image. Also, the pectoral removal takes place(Fig.3) and final figure shows that the whether the tumor is malignant or benign in Fig.4. Detect the classification using neural network.

3. RESULT AND DISCUSSION

The perform mammogram classification. Each approach has implemented using Matlab and evaluated for the efficiency of the classification. For the evaluation, it has used different data sets with more number of samples. In order to evaluate the proposed method, Mammographic Image Analysis Society (MIAS) data set has developed.

Table 1: Usage of database

Database	Number of samples	Number of testing images
MIAS	322	12



PI: Pectoral Image, EI: Extraction Image, PR: Pectoral Removal Muscle, A: Assessment.

Fig.4. Detect the classification using Neural Network

.

Table: 2. Time complexity

Image	Time in seconds	Over all Time Complexity
IMG001	0.011259	0.017799
IMG002	0.010111	
IMG003	0.011304	
IMG004	0.026614	
IMG005	0.026870	
IMG006	0.025379	
IMG007	0.017958	
IMG008	0.020682	
IMG009	0.011040	
IMG010	0.011230	
IMG011	0.020185	
IMG012	0.020959	

The table 2 shows time complexity which is obtained from 12 images. The second column shows time in seconds for 12 different images. The third column shows the overall time complexity which is obtained as 0. 017799.The neural network has used 10 layers in Fig.5.It is best performance 9.084e-06 at epoch 103in Fig.6.

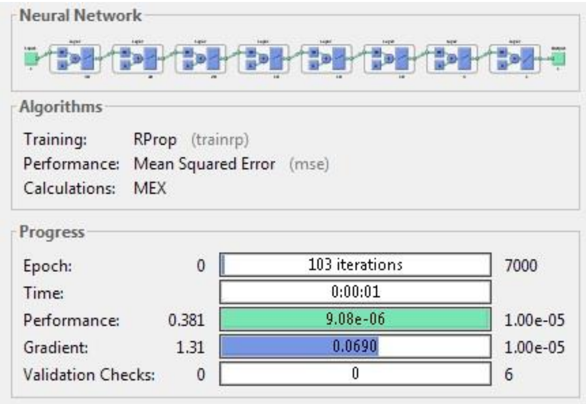


Fig. 5. Neural Network training.

Fig. 6. Best Training Performance

In [18] Let g be a gradient. Then g is a vector with n components that is defined for any point of a (differential) n-dimensional function f(x1, x2, . . . , xn). The gradient operator notation is defined as

$$g(x1, x2, \ldots, xn) = \nabla f(x1, x2, \ldots, xn) \qquad (4)$$

g directs from any point of f towards the steepest ascent from this point, with |g| corresponding to the degree of this ascent.Let f be an n-dimensional function ands = (s1, s2, . . . , sn) the given starting point. Gradient descent means going from f(s) against the direction of g, i.e. towards −g with steps of the size of |g| towards smaller and smaller values of f. According to it has shown fig.7.

Confirmatory tests were conducted with the same experimental setup to validate the accuracy of the results obtained. The results of confirmatory tests are presented in Fig.8.

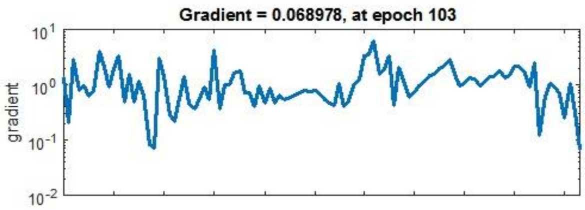


Fig. 7. Gradient of the Neural Network

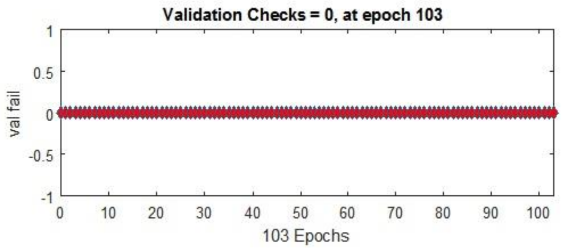


Fig. 8. Validation of the Neural Network

The accuracy of each proposed techniques were shown in percentage and neatly represented in graphical manner as shown in fig 9. Accuracy is the ratio between the numbers of true positive results to that of the total number of samples have been taken. The accuracy of MIAS method is of 98.4472%.

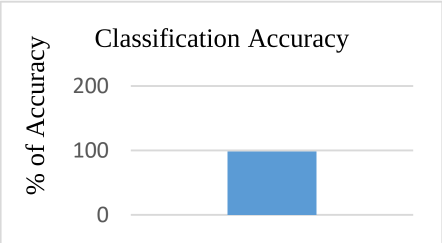


Fig. 9.NN of classification accuracy.

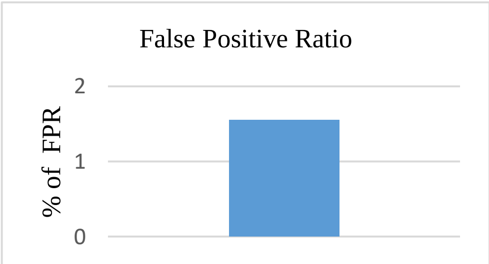


Fig.10. NN of false positives classification ratio

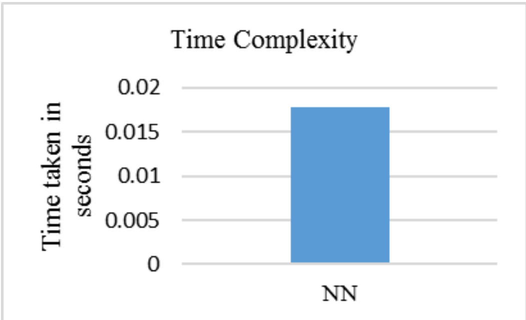


Fig.11. NN of Time Complexity

The false positive ratio is defined as the relation between numbers of false positive result to the total number of samples. In other words, it is the percentage difference from the accuracy which is having been shown in fig.10. MIAS method is of 1.552795%

Time complexity is the difference between the time classified and submitted. In this portion, the time requirement of each and every individual image have been calculated and they are averaged as shown in fig.11. By calculating under this method, the overall time complexity is 0.017799%

4. CONCLUSION

This paper deals with the computer-aided breast cancer identification based on the two different procedures such as the K-means algorithm and neural network. The K-means algorithm shows the tumor detected area result obtained from 12 different sample images. The data feature is obtained through entropy, mean, standard deviation, variance, co-variance and tumor detected area in sq.mm. On the whole, the second method used pectoral removal muscle and reduced the time complexity of about 0.017799% which is very low and concentrated false positive ratio of the neural network as well as the accuracy of the result when compared with the other computer-aided classification methods on breast cancer. In the future will be implemented in clusters in the thermography images.

REFERENCES

[1] M. Bazzocchi and F. Mazzarella,(2006) “CAD systems for mammography: a real opportunity”A review of the literature,” http://www.springerlink.com/content/x3157r8u72196h45/fulltext.pdf/,.

[2] A. Oliver,(2006) “A new approach to the classification of mammographic masses and normal breast tissue,” in Proceedings of the 18th International Conference on Pattern Recognition, 2006,vol. 4, pp. 707 – 710.

[3] J. Wei, H-P. Chan, B. Sahiner, C. Zhou, and L. M. Hadjiiski,(2009) “Computer-aided detectionof breast masses on mammograms: Dual system approach with two-view analysis,” Med.Phys., vol. 36, no. 10, pp. 4451 – 4460.

[4] Gou, S., Yang, H., Jiao, L., Zhuang, X. (2010): Algorithm of Partition Based Network Boosting for Imbalanced Data Classification. The International Joint Conference on Neural Networks (IJCNN).1-6. IEEE.

[5] Heang-Ping Chan, Jun Wei, Tao Wu and Mark A. Helvie; (2005) "Computer-aided Detection System for Breast Masses on Digital Tomosynthesis Mammograms: Preliminary Experience", Radiology, Vol. 237, No. 3.

[6] Nicholas Petrick, Berkman Sahiner and Lubomir M. Hadjiiski;(2002) "Breast Cancer Detection: Evaluation of a Mass-Detection Algorithm for Computer- aided Diagnosis Experience in 263 Patients", Radiology, Vol. 224, No. 1.

[7] Reyer Zwiggelaar, James E. Schumm and Christopher J. Taylor; "The Detection of Abnormal Masses in Mammograms", Wolfson Image Analysis Unit, University of Manchester, Manchester, UK.

[8] Jun Liu, Xiaomig Liu, Jianxun Chen and J Tang.(2011) Improved Local Binary Patterns for

Classification of Masses using Mammography. IEEE proceedings; 2692-2695.

[9] Muhammad Hussain, Summrina Kanwal Wajid, Ali Elzaart and Mohammed Berbar.(2011)A Comparison of SVM Kernel Functions Breast cancer Detection, Eighth International Conference computer Graphics, Imaging and Visualization IEEE.

[10] Fatima Eddaoudi, Fakhita Regrgui, Abdelhak Mahmoudi and Najib Lamouri.(2011)Mass Detection Using SVM Classifier Based on Texture Analysis. Applied Mathematical Science; 5: 367-379

[11] Leonardo de Oliveira Martins, Geraldo Braz Junior, Aristofanes Correa Silva, Anselmo Cardoso de Paiva and Marcelo Gattass. (2009) Detection of Masses in Digital Mammograms Using K-Means and Support Vector Machine. Electronics letters on Computer Vision and Image Analysis; 8(2); 39-50.

[12] B.N.Beena Ullala Mata and Dr.M.Meenakshi (2011).A Novel Approach for Automatic detection of Abnormalities in Mammograms, IEEE proceedings ; 831-836.

[13] Nalini Singh, Ambarish G Mohapatra and GurukalyanKanungo(2011). Breast Cancer Mass Detection in Mammograms Using K-means and Fuzzy C-means Clustering., International Journal of Computer Applications; 22: 15-21.

[14] B.N.Beena Ullala Mata and Dr.M.Meenakshi .(2011) A Novel Approach for Automatic detection of Abnormalities in Mammograms, IEEE proceedings ; 831-836.

[15] Kemal Polat and Sahil Gunes.(2007)," Breast Cancer Diagnosis using Least Square Support Vector machine". Digital Signal Processing; 17: 694-701.

[16] Christos Stergiou and Dimitrios Siganos. Neural Network,Available onhttp://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html.

[17] http://www.dkriesel.com/en/science/neural_networks

[18] S.Julian Savari Antony, Dr.S.Ravi,(2015) "Detection Of Masses in Digital Mammograms Using K Means And Neural Network",Published of Research India Publications, International Journal of Applied Engineering Research (IJAER),Volume 10, Number 7 pp. 17643-17656,ISSN: 09739769, 09734562.